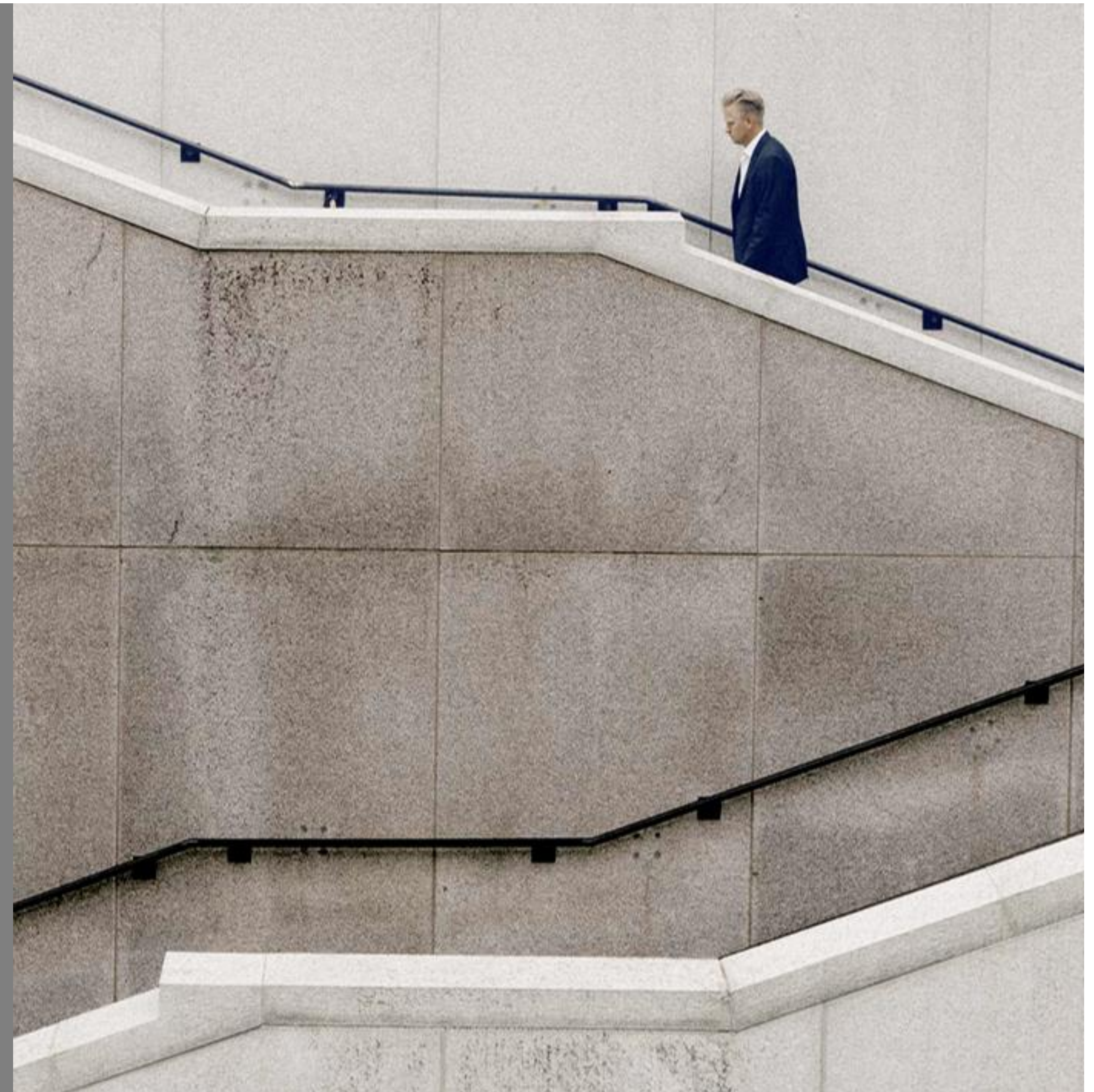


CORONATHON

EQUIPE 28

**PROPOSTA DE API PÚBLICA
BASEADA EM MODELAGEM
ESTATÍSTICA PARA
MELHORIA DE
ASSERTIVIDADE DO SINE**

*Participantes: Allan Rogge, Érico
Encarnação, Ivan Ormenesse, Jonas Lima,
Maurício Trujillo, Rodrigo Figueira*



Roadmap de Desenvolvimento

IMERSÃO

Avaliação dos principais problemas apresentados pelo time do SINE



IDEAÇÃO

Avaliação os problemas que vamos solucionar com base em nossas habilidades e informações disponíveis



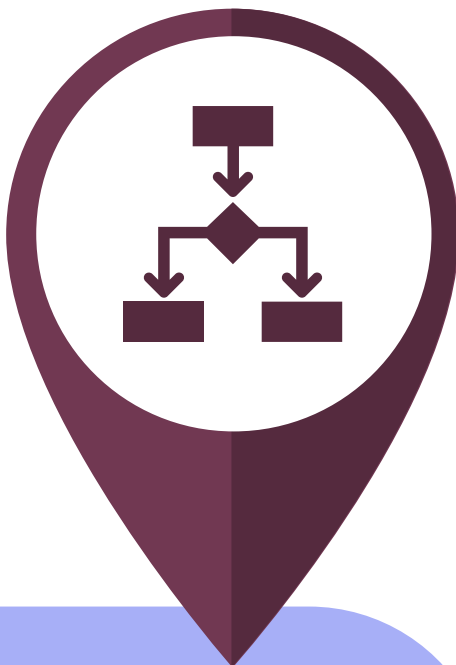
PROTOTIPAÇÃO

Desenhamos a solução ideal técnica ideal

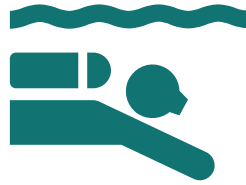


SOLUÇÃO

Entregamos um protótipo e sugestão de roll-out via api e algumas implicações para o front end

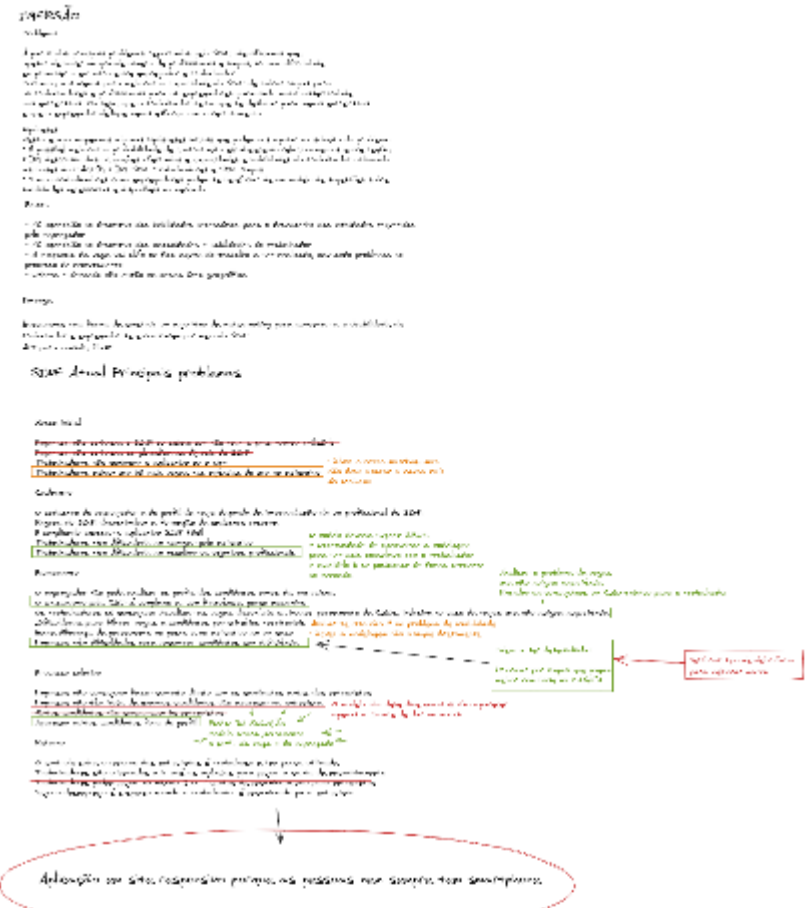
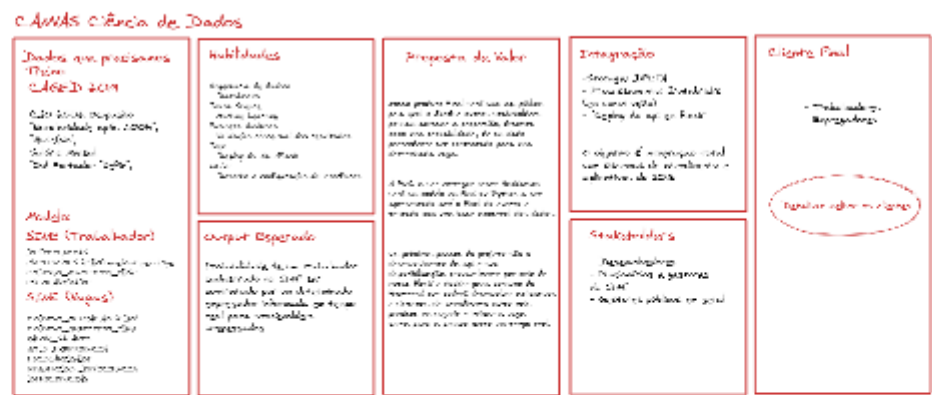
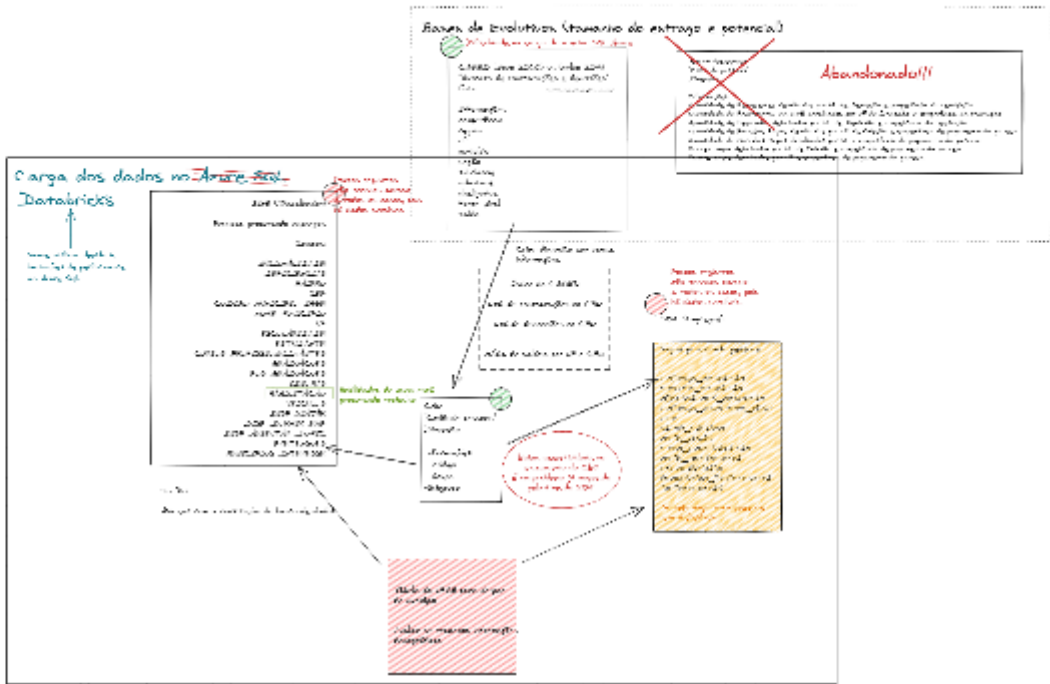


Imersão

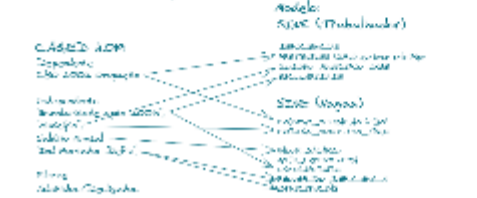


Iniciamos o entendimento do problema discutindo detalhadamente os principais problemas apresentados pelo SINE e, à partir da explicação das bases apresentadas nas primeiras etapas do evento, escolhemos alguns problemas para resolver

Processo criativo



Desenho Variáveis para ETL



Problemas a serem solucionados

Busca inicial

Trabalhadores têm dificuldade em escolher os objetivos profissionais

Pareamento

O cruzamento pela CBO é complexo e, com frequência, pouco assertivo

Os trabalhadores só conseguem visualizar as vagas disponíveis se houver pareamento de CBOs, inclusive no caso de vagas que não exigem experiência

Empresas têm dificuldades para encontrar candidatos com deficiência

Processo seletivo

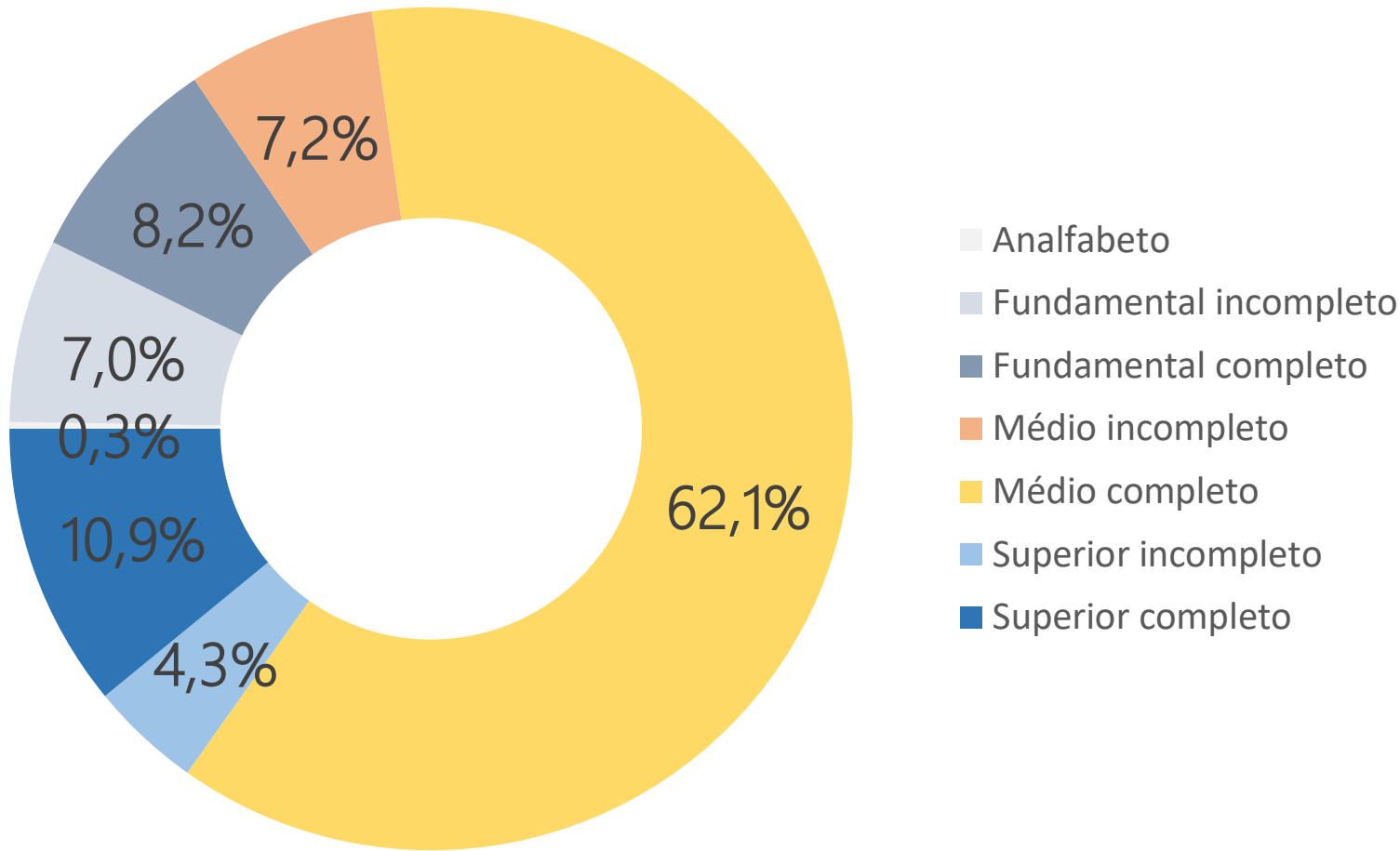
Aparecem muitos candidatos fora do perfil

Imersão – Análise Descritiva

Para a manipulação dos dados filtramos somente o Estado de São Paulo para admitidos (total de 3,7MM de registros) e identificamos que escolaridade e CBO tem forte impacto no salário do emprego. Daí surge a hipótese de que podemos melhorar a capacidade da ferramenta de indicar vagas e trabalhadores com base nestas informações

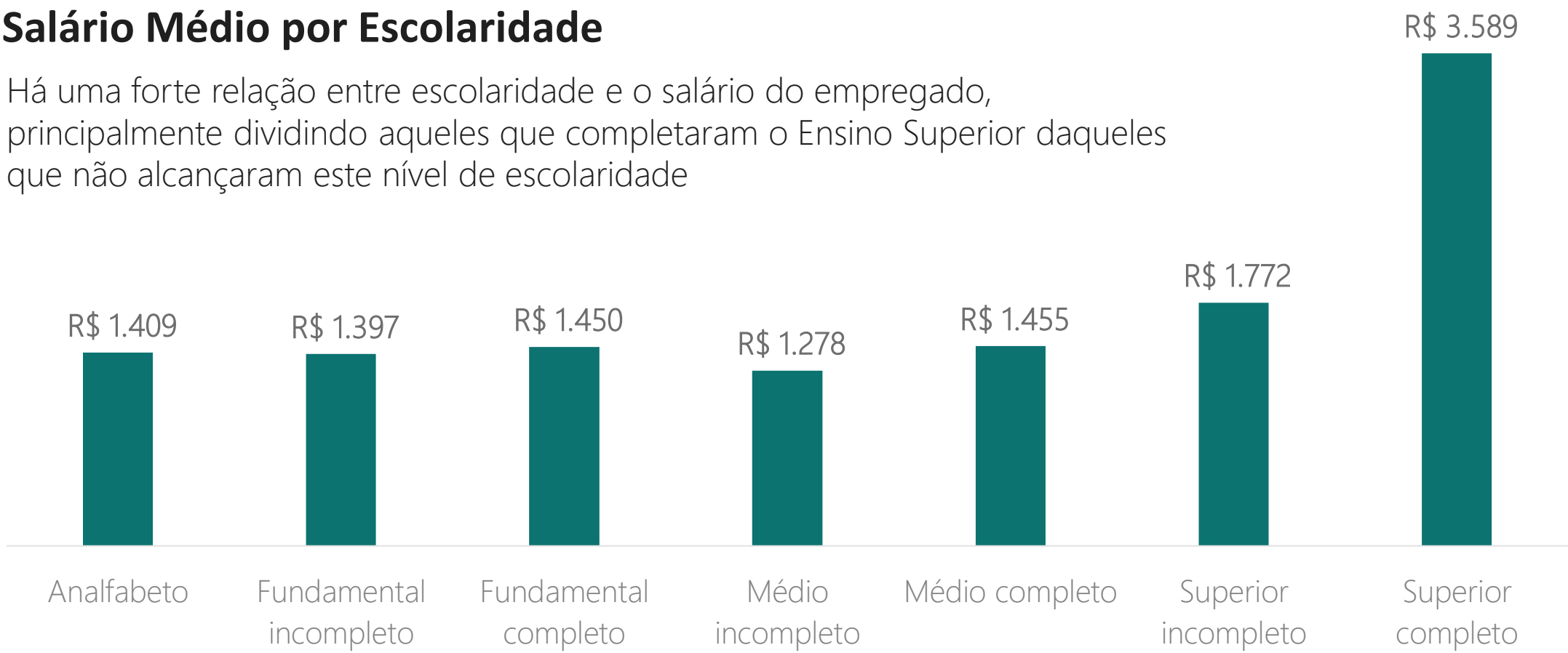
Análise por Escolaridade

A maior parte dos empregos cadastrados são de pessoas com ensino médio completo



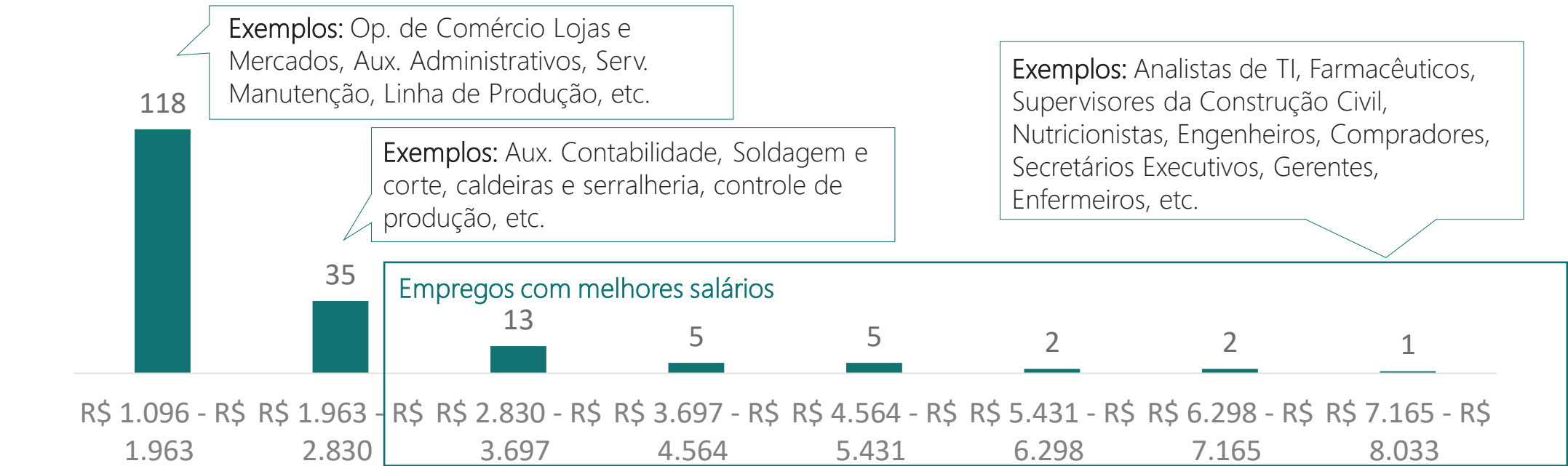
Salário Médio por Escolaridade

Há uma forte relação entre escolaridade e o salário do empregado, principalmente dividindo aqueles que completaram o Ensino Superior daqueles que não alcançaram este nível de escolaridade



Salário Médio por Escolaridade

Também há uma concentração dos CBO's em um nível relativamente baixo de remuneração indicando que pode haver perfis muito diferentes entre os grupos de pessoas



Ideação



À partir de uma extensa análise das bases definimos o desenho geral de uma solução:

Produto Final

Nosso produto final será uma API pública pela qual o SINE e, eventualmente, outros stakeholders possam acessar a propensão, descrita como uma probabilidade de um trabalhador ser contratado para uma determinada vaga. **Nossa missão é reduzindo o atrito e agilizar o processo de contratação para trabalhadores e empregadores**, melhorando a visualização de vagas no front-end e facilitando encontrar pessoas e posições, inclusive aquelas portadoras de deficiência.

Resultados da Etapa

Ideias adotadas

- _Modelo preditivo por área geográfica e ocupação para a predição à partir da base do CAGED
- _Utilização do Databricks versão comunidade para lidar com o volume de informações
- _A entrega final seria uma API que consulta um algoritmo de priorização de vagas e entrega resultado para trabalhadores, empregadores e stakeholders

Ideias abandonadas

- _Informações sobre o seguro desemprego, por serem dados agregados não contribuem para construção do modelo
- _Base da RAIS: entendemos que o CAGED de 2019 possuía as informações necessárias para a modelagem
- _BASE CAGED 2020: Poucas informações e dados fortemente impactados pela Covid-19
- _Utilização das bases do SINE tanto de vagas como de trabalhadores para treinar um modelo de machine learning, pois, devido à presença de informações sensíveis, o volume de dados disponibilizados é insuficiente

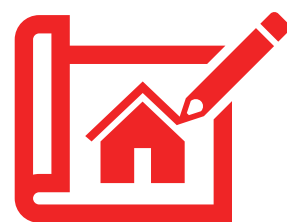


Ideação - CANVAS

O processo de ideação foi feito seguindo o seguinte CANVAS com o objetivo de direcionar a prototipação e desenvolvimento da solução

<div>Dados que precisamos</div> <div>Base de treino</div> <div>CAGED 2019</div> <div>CBO 2002 Ocupação</div> <div>"Escolaridade após 2005",</div> <div>"Município",</div> <div>Salário Mensal</div> <div>"Ind Portador Defic",</div> <div>Modelo</div> <div>SINE (trabalhador)</div> <div>Deficiências</div> <div>Pretensões(CBO) quebrar pelo pipe</div> <div>Codigo_municipio_ibge</div> <div>Escolaridade</div> <div>SINE (vagas)</div> <div>Codigo_ocupacao (cbo)</div> <div>Codigo_municipio_ibge</div> <div>Valor_salario</div> <div>Qtd_experiencia</div> <div>Escolaridade</div> <div>Requerido_deficiencia</div> <div>Deficiências</div>	<div>Habilidades</div> <div>Engenharia de dados:</div> <div>- Databricks</div> <div>Data Science</div> <div>- Machine Learning</div> <div>Business Analytics</div> <div>- Validação conceitual dos resultados</div> <div>Dev.</div> <div>- Deploy da Api /Flask</div> <div>UX/UI</div> <div>- Desenho e configuração de interfaces</div> <div>Output Esperado</div> <div>Probabilidade de um trabalhador cadastrado no SINE ser contratado por um determinado empregador informada em tempo real para stakeholders Interessados.</div>	<div>Proposta de valor</div> <div>Nosso produto final será uma API pública pela qual o SINE e outros stakeholders possam acessar a propensão, descrita como uma probabilidade, de um dado pretendente ser contratado para uma determinada vaga.</div> <div>A POC a ser entregue neste Hackathon será um modelo em Rmd ou iPython a será apresentado até o final do evento e testado com uma base amostral dos dados.</div> <div>Os próximos passos do projeto são o desenvolvimento da API e sua disponibilização, provavelmente por meio de Flask para consumo do front-end que poderá desenvolver aplicativos e sistemas de atendimento muito mais precisos em sugerir e indicar a vaga certa para a pessoa certa em tempo real.</div>	<div>Integração</div> <div>-Storage: DELTA</div> <div>- Processamento: Databricks</div> <div>- Deploy da api em flask</div> <div>O objetivo é integração total com sistemas de atendimento e aplicativos do SINE (Trabalha Brasil)</div> <div>Stakeholder</div> <div>- Desenvolvedores</div> <div>- Funcionários e gestores do SINE</div> <div>- Gestores públicos em geral</div>	<div>Cliente Final</div> <div>Em última instância a solução tem por propósito atender a necessidade tanto de trabalhadores como de empregados:</div> <div>- Trabalhadores que buscam emprego na plataforma Trabalha Brasil ou em uma agência do SINE</div> <div>- Empregadores que oferecem vagas por meio da plataforma do Trabalha Brasil</div>
---	--	--	---	---

Prototipação

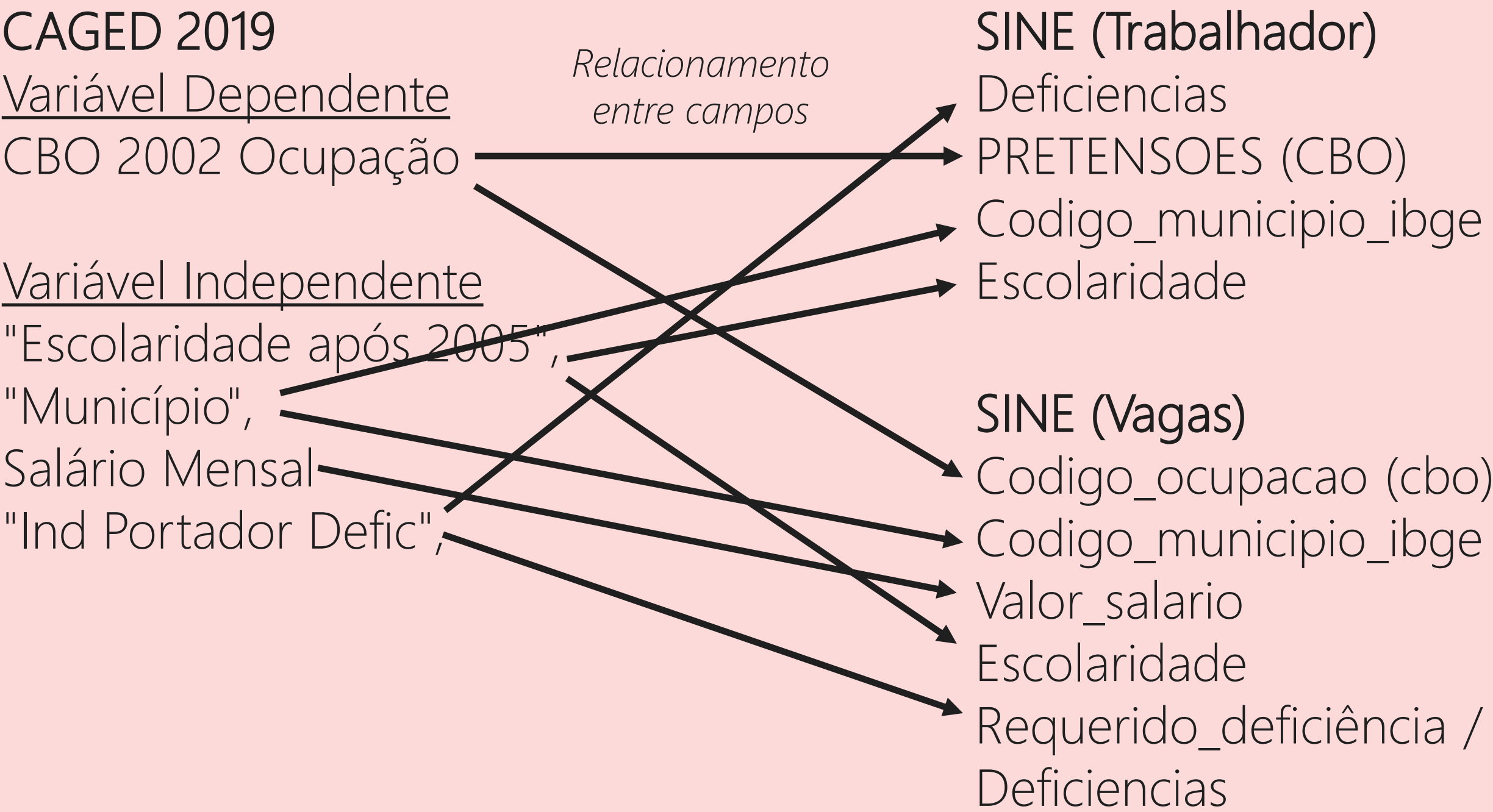


Analizamos em detalhes as bases do SINE e da CAGED pra entender especificamente quais informações poderiam ser utilizadas na modelagem, como trata-las e como aplicar modelos estatísticos à elas.

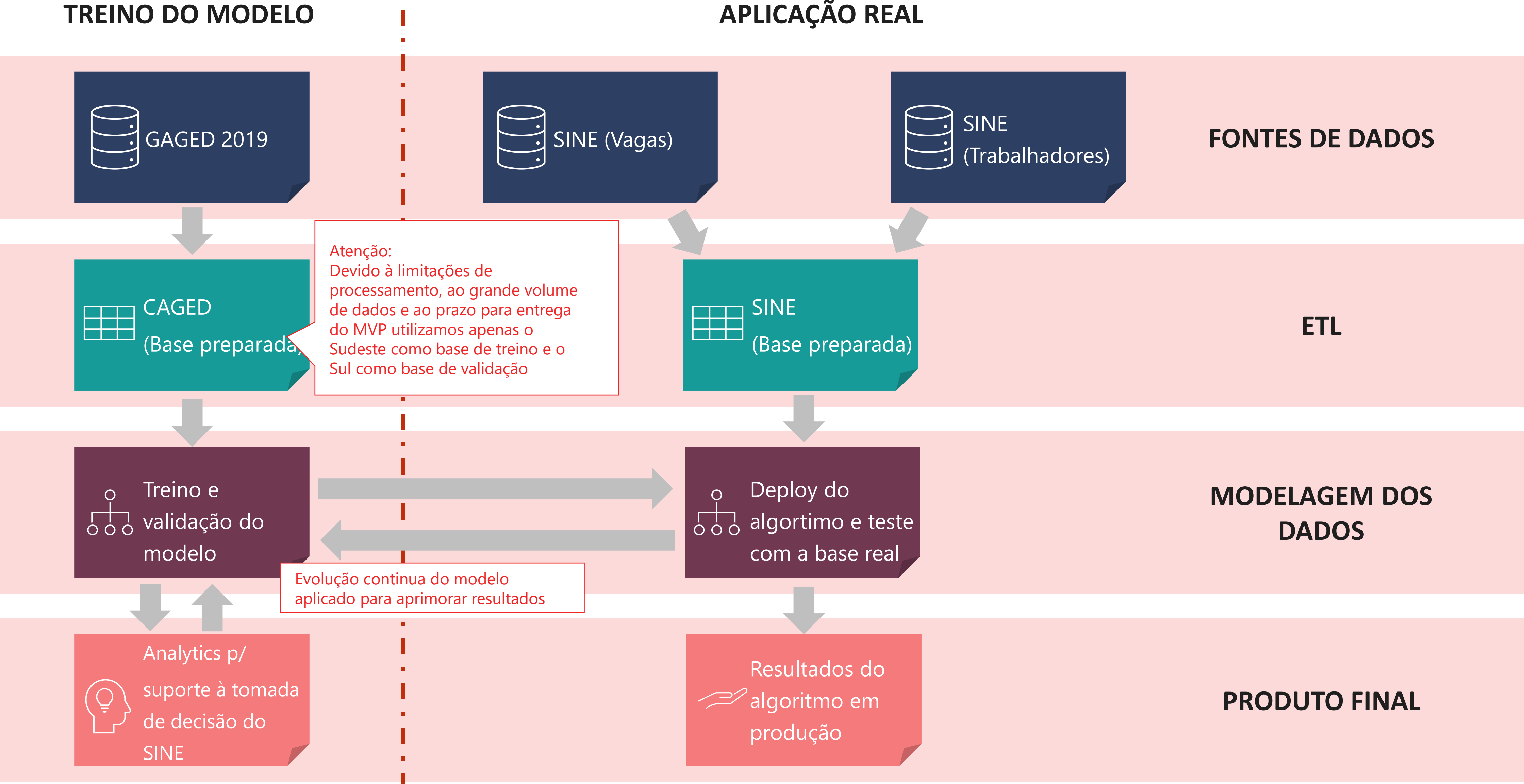
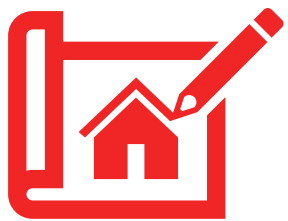
Desenhamos o processo de ETL e selecionamos o Databricks versão comunidade como aplicação para carga e modelagem dos dados.



Descritivo do relacionamento entre variáveis



Prototipação – Estrutura de Dados e desenho da solução



Prototipação - Modelo



Aplicou-se o modelo de Regressão Logística para cada CBO da base de demitidos e admitidos do CAJED, para a região de São Paulo. Foram selecionados 30 CBOs para um primeiro experimento, devido capacidade de servidor e espaço para treinamento do modelo.

Foram gerados 30 modelos, um para cada CBO, com o objetivo de descobrir a probabilidade que cada trabalhador possui em relação ao CBO que ele se identifica. Com isso, foi calculada a acurácia de cada modelo, escolhendo como critério a probabilidade de 0,2 para o sucesso da ocorrência.

As acurácias oscilaram e conseguiram alcançar poder de predição de até 95%. Deseja-se melhorar o ajuste do modelo a fim de encontrar os melhores pontos de cortes para cada CBO e consequentemente melhorar o poder de predição.

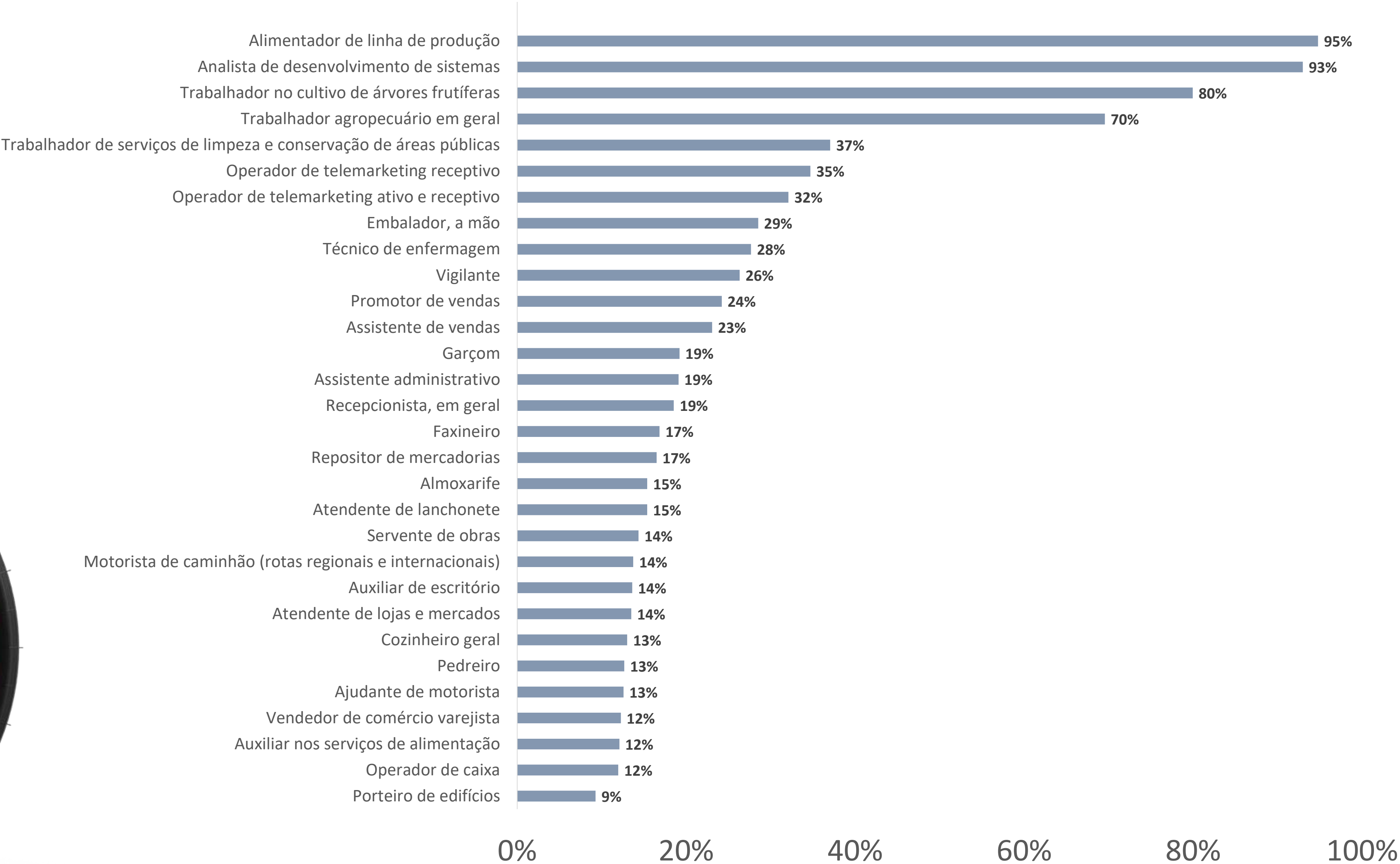
Além disso, deseja-se aplicar o modelo para a base nacional do CAGED e para a base de indivíduos que procuram vagas de emprego do SENE.



Prototipação - Modelo



Acurácia do modelo de acordo com CBO



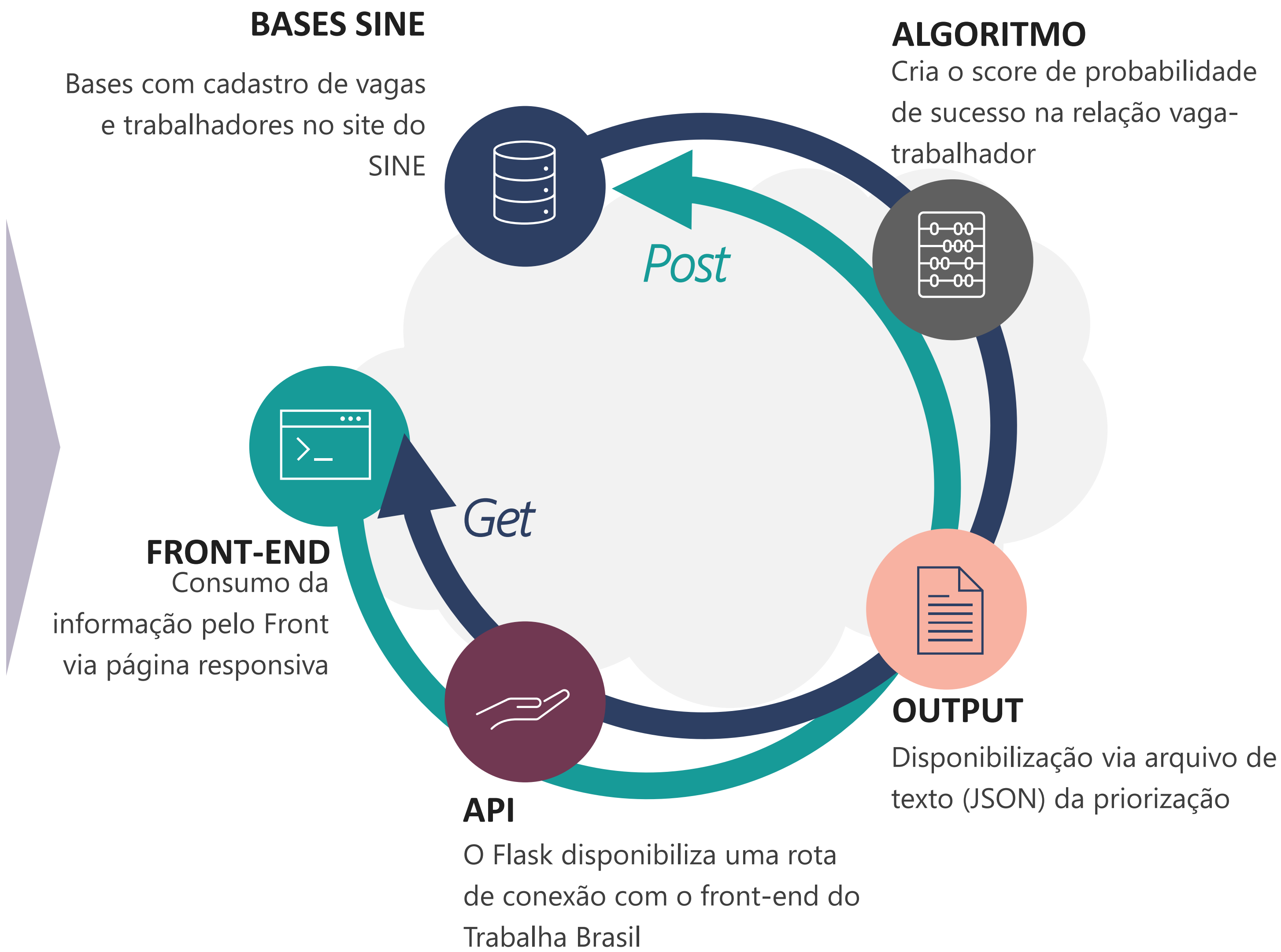
Solução – Disponibilização via Flask

Descritivo Conceitual da Aplicação

Nesta abordagem, o Flask irá realizar um túnel de API para conectar a base de dados ao nosso modelo e ao front-end do Trabalha Brasil.

Ao realizar o cadastro, será feito um post em direção a base, para inserir os dados via JSON na base de dados.

Ao solicitar uma procura de emprego, um método 'get' consome a informação do modelo e outro 'get' exhibe a informação na tela



Solução – Aplicação e Deploy

O portal Trabalha Brasil é bastante claro e intuitivo, tanto para o Trabalhador como para um atendente em uma agência do SINE...

Versão Desktop

TRABALHA BRASIL

Sou TrabalhadorSou Empregador

Olá, Mayara!Saír

Busca de vagas

Buscar vagas por Área

Buscar vagas por Estado

Alterar Currículo

Alerta de Vaga

Média Salarial

Função ou Palavra Chave

Cidade/Estado (ex. São Paulo/SP)

🔍 BUSCAR VAGAS

Trabalha Brasil / Cadastrar Currículo

CADASTRE SEU CURRÍCULO GRATUITAMENTE

Informe os dados abaixo e tenha acesso a milhares de vagas atualizadas diariamente

1Mini Currículo

2Experiência e Escolaridade

CPF

999.999.999-00

Nome

Joana Hernandez

Email

hernandez.jo@email.com

Nível de Formação

Superior Completo

Cidade e Estado

São Paulo / SP

Sexo

☐ Masculino

☒ Feminino

Posso ter deficiência

☐ Sim

☒ Não

Data de Nascimento

23/11/1994

Celular

(11) 99999-0000

Telefone

(11) 5000-0000

Função

Designer

Proteção Salarial

5.000,00

Bairro

Pinheiros

☒ Concorde com a política de privacidade

SALVAR E CONTINUAR



Versão Mobile

TRABALHA BRASIL

Função ou Palavra Chave

Cidade/Estado (ex. São Paulo/SP)

🔍 BUSCAR VAGAS

CADASTRE SEU CURRÍCULO GRATUITAMENTE

Informe os dados abaixo e tenha acesso a milhares de vagas atualizadas diariamente

1Mini Currículo

2Experiência e Escolaridade

CPF

999.999.999-00

Nome

Joana Hernandez

Email

hernandez.jo@email.com

Sexo

☐ Masculino

☒ Feminino

Posso ter deficiência

☐ Sim

☒ Não

Data de Nascimento

23/11/1994

Celular

(11) 99999-0000

Telefone

(11) 5000-0000

Função

Designer

Proteção Salarial

5.000,00

Cidade e Estado

São Paulo / SP

Bairro

Pinheiros

☒ Concorde com a política de privacidade

SALVAR E CONTINUAR

Solução – Aplicação e Deploy

Portanto, a utilização da API atuaria principalmente no back-end, sem muitas alterações no front-end melhorando a recomendação de vagas que a ferramenta já faz atualmente

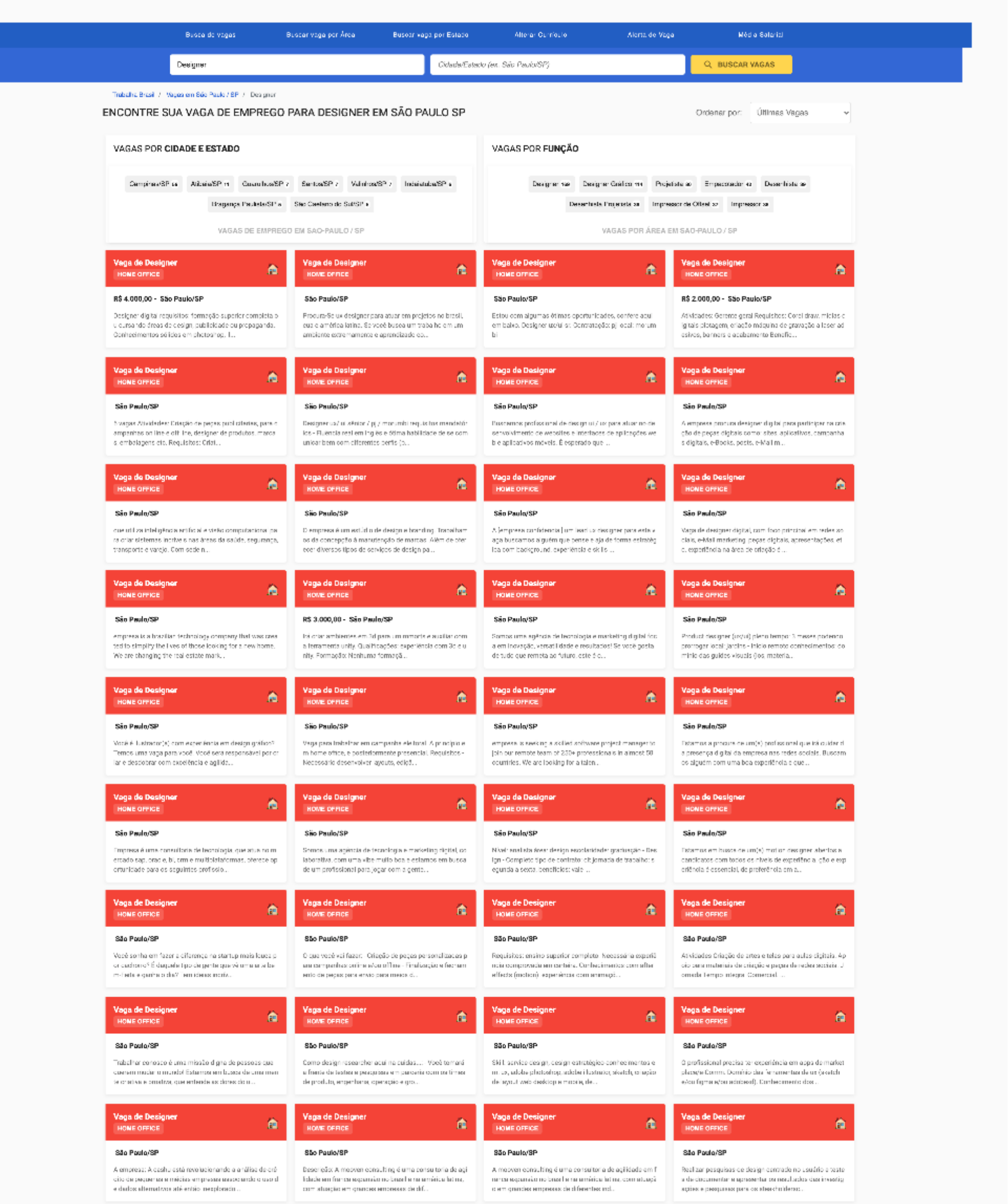
Novas Features

— O trabalhador pode filtrar vagas por cidade e estado de forma mais facilitada com os filtros no início da página ao invés de no final da página

— Consultado a API as vagas mais relevantes, ou seja, aquelas com maior probabilidade de contratação serão exibidas na página

— O empregador também pode consultar os currículos com base na mesma relevância que o trabalhador, reduzindo atritos e a quantidade de entrevistas improdutivas

Melhorias na página de busca de empregos



A Equipe



Allan Rogge

DBM na MarketData

Estudante de Análise e desenvolvimento de sistemas atua na área de dados e busca migrar para a área de Ciência de Dados



Érico Luiz Encarnação Rocha

Consultor de CX Analytics na Vivo

Profissional com mais de 9 anos de experiência na área de tecnologia atuando com BI, Analytics, desenvolvimento de software, liderança e gestão de projetos de TI.



Ivan Ormenesse

Head of Data Analytics na Juntos Somos Mais

Mais de 10 anos de experiência em pesquisa de mercado, customer insights, inteligência de mercado e sólido conhecimento em estratégias de dados em setores diversos.



Jonas Pinheiro de Lima

DBM Sênior na Juntos Somos Mais

Possui experiência em programação baseada na Web, e conhecimento estatístico, busca seguir carreira na área de Data Science.



Mauricio Persinotti Trujillo

Analista de Data Analytics na Juntos Somos Mais

Profissional com forte perfil analítico, boa visão holística dos processos e interações do negócio e facilidade de transição entre os níveis operacional, tático e estratégico.



Rodrigo Almeida

People Analytics na DASA

Bacharel em Estatística pela UFPA. Atuou como em diversos setores executando modelos preditivos para a melhorias de processos e iniciativas de negócios.